Fermi National Accelerator Laboratory

FERMILAB-Pub-97/410-A

Likelihood Analysis of Galaxy Surveys

Scott Dodelson, Lam Hui and Andrew Jaffe

Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510

December 1997

Submitted to *Astrophysical Journal*

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CH03000 with the United States Department of Energy

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Distribution

Approved for public release; further dissemination unlimited.

Fermi National Accelerator Laboratory

FERMILAB–PUB–97/97-410-A
astro-ph/9712074
December 9, 1997

Likelihood Analysis of Galaxy Surveys

Scott Dodelson[*]
NASA/Fermilab Astrophysics Center
Fermi National Accelerator Laboratory, Batavia, IL 60510

Lam Hui[†]
NASA/Fermilab Astrophysics Center
Fermi National Accelerator Laboratory, Batavia, IL 60510

Andrew Jaffe[‡]
Center for Particle Astrophysics, 301 LeConte Hall
University of California,
Berkeley, CA 94720

One of the major goals of cosmological observations is to test theories of structure formation. The most straightforward way to carry out such tests is to compute the likelihood function $\mathcal{L}$, the probability of getting the data given the theory. We write down this function for a general galaxy survey. The full likelihood function is very complex, depending on all of the n-point functions of the theory under consideration. Even in the simplest case, where only the two point function is non-vanishing (Gaussian perturbations), $\mathcal{L}$ cannot be calculated exactly, primarily because of the Poisson nature of the galaxy distribution. Here we expand $\mathcal{L}$ about the (trivial) zero correlation limit. As a first application, we take the binned values of the two point function as free parameters and show that $\mathcal{L}$ peaks at $(DD - DR + RR)/DD$. Using Monte Carlo techniques, we compare this estimator with the traditional DD/DR and Landy & Szalay estimators. More generally, the success of this expansion should pave the way for further applications of the likelihood function.

[*] Electronic mail: `dodelson@hermes.fnal.gov`
[†] Electronic mail: `lhui@fnal.gov`
[‡] Electronic mail: `jaffe@physics.berkeley.edu`

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

I. INTRODUCTION

Recently, there has been a vast increase in both the quality and quantity of data accumulated by observational cosmologists. The high quality of this observational work sets a standard for theorists to interpret the data as carefully and meaningfully as possible. For these purposes, the likelihood function has become a powerful tool for all cosmologists. The likelihood function, $\mathcal{L}$, is the probability of a given data set given a theory; as such it is a natural way of connecting observations with theory. Examples of its application to cosmological data sets include cosmic microwave background anisotropies, gravitational lens studies and fitting absorption lines in QSO spectra.

Until now, though, there has been no attempt to apply likelihood techniques to galaxy surveys. Instead, one typically determines the two-point function, $\xi(r)$, or the power spectrum, $P(k)$, from the survey using appropriately chosen estimators (usually depending on the distribution of observed pairs of points). On the one hand, this is quite surprising: it would be wonderful to use all the information in a survey instead of just one small subset of it. If the galaxy distribution was Gaussian, so that only the two-point function mattered, it would make sense to neglect all the other information in the catalogue. The galaxy distribution, however, is decidedly non-Gaussian, so the likelihood function, which incorporates information from all the n-point functions, seems more appropriate. On the other hand, the likelihood function for galaxies in a survey is a very complicated beast. Even if the underlying density field were Gaussian (so that only the two point function was non-zero), we would not necessarily expect an ad hoc estimator to be ideal. Further, and perhaps most importantly, the galaxies themselves are a Poisson (or other) sample of this underlying field, which introduces non-Gaussianity into the distribution of galaxy locations. In §II, we write down $\mathcal{L}$ and show that even for the Gaussian case it is computationally unfeasible to calculate.

In §III, we propose one way to extract some information from the likelihood function: we expand $\mathcal{L}$ about its value when all correlations are zero. This weak correlation limit proves very fruitful. As an example, we consider a general 2-D (angular) survey (although the validity of our results is not confined to 2-D) and a theory with the free parameters being the value of the two-point correlation function w_θ in different angular bins. Thus, in this example, $\mathcal{L}$ is a function of the data and the parameters w_θ. We show that, for each θ, $\mathcal{L}$ peaks at

$$w_\theta^{\mathcal{L}} = \frac{DD(\theta) - DR(\theta) + RR(\theta)}{DD(\theta)} \tag{1}$$

where $DD, DR,$ and RR are respectively the number of pairs of particles in the data set with angular distances within the bin denoted by θ; the number of pairs of particles—one in the data set and one in a random catalogue—in the θ bin; and the number of random pairs in the bin θ. This estimator is very close to the one introduced by Landy & Szalay (1993). (For other estimators, see Peebles (1980), Hewett (1982), and Hamilton (1993); for a recent discussion of the Landy & Szalay estimator and a generalization to higher order correlations, see Szapudi & Szalay (1997)). We regard this as a success of the expansion: we are able to extract a reasonable estimator, one shown to be significantly more effective than the traditional DD/DR. In §IV, we analyze these estimators more carefully using Monte Carlo simulations to see which has the lowest variance. These simulations suggest that the maximum likelihood estimator of Eq. 1 has a smaller variance than both the traditional estimators and the Landy & Szalay estimator. In the course of performing these simulations, we have uncovered additional terms in the variance of the Landy-Szalay estimator which dominates over the normal Poisson term when there are many galaxies in the survey. We present a simple derivation of these additional terms in Appendix B.

While the application developed in §III and §IV is useful, we feel the most important feature of our analysis is the realization that the likelihood function can be approximated in a meaningful way. This opens the door to a host of applications, some of which we speculate about in the conclusion. Finally, to improve the readability of the text, we have shifted most of the calculational details of the expansion in §III to Appendix A.

II. The Likelihood Function

The probability of finding N galaxies at positions $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N$ and no galaxies elsewhere in the survey volume V is given by

$$\begin{aligned} \mathcal{L} &\equiv P[\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N; \Phi_0(V)|w_2(\theta), n] = \exp\{W_0\}\ n^N dV_1 dV_2 \cdots dV_N \\ &\times \Bigg[W_1(\vec{x}_1) W_1(\vec{x}_2) \cdots W_1(\vec{x}_N) \\ &+ W_2(\vec{x}_1, \vec{x}_2) W_1(\vec{x}_3) \cdots W_1(\vec{x}_N) + \text{permutations} \\ &+ \ \ldots \end{aligned}$$

$$+ \quad W_2(\vec{x}_1, \vec{x}_2) \cdots W_2(\vec{x}_{N-1}, \vec{x}_N) + \text{permutations}\Big] \tag{2}$$

where $\Phi_0(V)$ is the "proposition" that there are no galaxies in the sample other than those at the specified $\vec{x}_i$. Here we have assumed that there are no higher order correlations. The general expression including higher order correlations was first written down by White (1979); we do not reproduce it here. So the free parameters in the "theory" are the density n and the two point correlation function, $w_2(\theta)$. The W_i's are then given simply by

$$\begin{aligned} W_0 &= -nV + \frac{n^2}{2} \int dV_1 dV_2 w_2(\vec{x}_1, \vec{x}_2) \\ W_1(\vec{x}_1) &= 1 - n \int dV_2 w_2(\vec{x}_1, \vec{x}_2) \\ W_2(\vec{x}_1, \vec{x}_2) &= w_2(\vec{x}_1, \vec{x}_2) \end{aligned} \tag{3}$$

Here the infinitesmal "volume" dV_i surrounds the position $\vec{x}_i$ and can be either 2-D (areas) or 3-D (volumes). For a realistic survey, all of these volume integrals are to be weighted by the selection function of the survey: this replaces the probability that there are galaxies at the $\vec{x}_i$ and no others with the probability of *observing* galaxies at the $\vec{x}_i$ and nowhere else. In principle, by suitably recomputing the probability of the statement $\Phi_0(V)$, complications such as redshift-space distortions could be included.

We should also note that this expression assumes a particular model for the relationship between galaxy positions and the underlying "number density field": the galaxies are a *Poisson sample* of the density. That is, the probability of finding a single galaxy in an infinitesimal volume δV around $\vec{x}$ is just $n(\vec{x})\delta V \ll 1$. We can then use a suitable biasing prescription to connect the number density field $n(\vec{x})$ to the mass density field.

Equation 2 is deceptively simple. A given line contains m occurences of the two point function W_2. The deceptiveness lies in the phrase "+ permutations": there are many, many terms included in this phrase. Consider the line with $N/8$ W_2's. The number of ways of choosing $3N/4$ galaxies and assigning each a factor of $W_1(\vec{x}_i)$ is

$$\frac{N!}{(3N/4)!\,(N/4)!}. \tag{4}$$

The remaining $N/4$ galaxies can be arranged into W_2's in

$$\frac{(N/4)!}{2^{N/8}\,(N/8)!} \tag{5}$$

ways. So the total number of terms on this one line is

$$\frac{N!\ 2^{-N/8}}{(N/8)!\ (3N/4)!} \simeq (8N)^{N/8} \tag{6}$$

where the approximate equality uses Stirling's formula. Even for a small catalogue with a thousand galaxies, this line contains 10^{488} terms! Each of these must be calculated separately; and there are $N/2$ lines with comparable numbers of terms (and this is just for the Gaussian case!). Clearly, an exact calculation of the likelihood function is out of the question. We need to develop approximation schemes that will enable us to circumvent an exact calculation.

III. Weak Correlation Limit

We are especially interested in the regime where the correlations are weak ($w \ll 1$). In this regime, we might further expect non-Gaussianity to be negligible, at least when we start from an initially Gaussian density field as from inflationary theories.

A. Zero Order Solution

We want to expand Eq. 2 about $w_2 = 0$. So let us first find the likelihood when $w_2 = 0$. In this case the likelihood function is extremely simple; it reduces to the Poisson distribution:

$$P[\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N; \Phi_0(V)] = \exp\{-nV\}\ n^N dV_1 dV_2 \cdots dV_N \tag{7}$$

If we differentiate this with respect to the one free parameter, the density n, we find that the likelihood function peaks when

$$n = \bar{n} = N/V. \tag{8}$$

The width of the likelihood function gives a measure of how accurately the parameter n is known. One way to estimate this width is to expand $\ln(\mathcal{L})$ around $n = \bar{n}$:

$$\ln(P) = \ln(P[\bar{n}]) + \frac{1}{2}\frac{d^2 \ln P}{dn^2}|_{n=\bar{n}}(n-\bar{n})^2 + \ldots \tag{9}$$

In this simple example,

$$\begin{aligned}\ln(P) \quad &= -nV + N\ln(n) + \text{constant} \\ &= -\bar{n}V + N\ln(\bar{n}) + \frac{1}{2}\left(\frac{-N}{\bar{n}^2}\right)(n-\bar{n})^2 + \ldots \end{aligned} \qquad (10)$$

So we can identify the width of the likelihood function as $(\bar{n}^2/N)^{1/2} = (N/V^2)^{1/2}$, so that $\delta n/n = N^{-1/2}$, which is of course the correct answer for a Poisson distribution. Note that the Poisson distribution considered as a function of n is skewed. So if we choose as our estimator (say) the mean rather than the mode, we would find a different answer; for a large survey, this is clearly a tiny effect.

We now pursue this approach including correlations. Specifically, we calculate the first derivative of $\ln(P)$ and set it to zero in order to determine the free parameters. Then we calculate the second derivative to calculate the width of the likelihood function.

B. First Order Solution for n

The starting point is Eq. 2:

$$\begin{aligned}\ln P = W_0 \quad &+ \quad N\ln n + \ln\left[\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} W_2(\vec{x}_a,\vec{x}_b)\prod_c^{a,b} W_1(\vec{x}_c)\right. \\ &+ \left.\frac{1}{2^3}\sum_a\sum_b^{a}\sum_c^{a,b}\sum_d^{a,b,c} W_2(\vec{x}_a,\vec{x}_b)W_2(\vec{x}_c,\vec{x}_d)\prod_e^{a,b,c,d} W_1(\vec{x}_e)\right]\end{aligned} \qquad (11)$$

where we have dropped irrelevant constants [such as $\ln V$]. We need to expand this to second order in w_2; then when differentiating to find the maxmium, we will get a linear equation for w_2. Therefore, only terms up to second order in W_2 have been kept. Another facet of this expansion—which is detailed in Appendix A—is that all the W_1's in the $W_2 W_2 \prod W_1$ term (last in the square brackets above) can be set to one. Similarly, all but one of the W_1's in terms of the type $W_2 \prod W_1$ can be set to one, etc. We have introduced the notation of superscripts on $\sum, \prod$. These indicate which galaxies should *not* be summed or mulitplied over. Thus, $\sum_a^b$ means sum over all $a = 1, \ldots, N$ except $a = b$. For example,

$$\sum_a^{b,c} \equiv \sum_a (1-\delta_{ab})(1-\delta_{ac}), \qquad (12)$$

the generalization to more or fewer superscripts should be plain.

Now, a word about the factors of 2, which may be a source of confusion: the term with one W_2 clearly has one factor of two to account for the fact that we are double counting $W_2(\vec{x}_1, \vec{x}_2)$ and $W_2(\vec{x}_2, \vec{x}_1)$. The term with two W_2's obviously needs two of these types of factors. But it also needs another one to account for $W_2(\vec{x}_1, \vec{x}_2)W_2(\vec{x}_3, \vec{x}_4)$ and $W_2(\vec{x}_3, \vec{x}_4)W_2(\vec{x}_1, \vec{x}_2)$, hence the $1/2^3$ factor before the last term.

It is worthwhile here to introduce the notation of Landy & Szalay. For these purposes we divide the survey volume into K cells, each of which is so small that it contains at most one galaxy. Then

$$\int dV = \frac{V}{K}\sum_{i=1}^{K} \tag{13}$$

where the sum over i includes even those cells that don't have galaxies in them. With this notation, W_0 and W_1 can be rewritten as

$$W_0 = -nV + \frac{n^2V^2}{2K^2}\sum_{i,j} w_2(\vec{x}_i, \vec{x}_j) \tag{14}$$

$$W_1(\vec{x}_1) = 1 - \frac{nV}{K}\sum_{i} w_2(\vec{x}_1, \vec{x}_i) \tag{15}$$

It is worth noting that if one goes over the original derivation by White (1979), the summation over i and j in the above two equations should, strictly speaking, range over only empty cells. We do not place such restrictions here, essentially working in the continuum limit, even though we represent the integrals as discrete sums.

Now let's find the value of the density at the maximum of the likelihood function. We need to differentiate Eq. 11 with respect to n.

$$\begin{aligned}\frac{\partial \ln P}{\partial n} &= \frac{\partial W_0}{\partial n} + \frac{N}{n} \\ &+ \left[\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^a W_2(\vec{x}_a, \vec{x}_b)\prod_c^{a,b} W_1(\vec{x}_c)\right. \\ &+ \left.\frac{1}{2^3}\sum_a \sum_b^a \sum_c^{a,b} \sum_d^{a,b,c} W_2(\vec{x}_a, \vec{x}_b)W_2(\vec{x}_c, \vec{x}_d)\right]^{-1} \\ &\times \frac{\partial}{\partial n}\left[\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^a W_2(\vec{x}_a, \vec{x}_b)\prod_c^{a,b} W_1(\vec{x}_c)\right]\end{aligned} \tag{16}$$

Here we have used the fact that $\partial W_2/\partial n = 0$. To go further we need the other partial derivatives:

$$\frac{\partial W_0}{\partial n} = -V + \frac{nV^2}{K^2}\sum_{i,j} w_2(\vec{x}_i, \vec{x}_j) \quad ; \quad \frac{\partial W_1(\vec{x}_a)}{\partial n} = -\frac{V}{K}\sum_i w_2(\vec{x}_i, \vec{x}_a) \tag{17}$$

Note first that setting Eq. 16 to zero will give a solution of the form $n = N/V + O(w_2)$. We therefore need to keep only terms linear in w_2. After differentiating the numerator, it will be of order w_2. We keep only these linear terms and set the denominator to one. Thus, we are left with

$$\begin{aligned}\frac{\partial \ln P}{\partial n} &= -V + \frac{N}{n} + \frac{nV^2}{K^2}\sum_{i,j} w_2(\vec{x}_i, \vec{x}_j) - \frac{V}{K}\sum_{a,i} w_2(\vec{x}_i, \vec{x}_a) \\ &= -V + \frac{N}{n} + \frac{V}{K}\sum_i \left[\frac{nV}{K}\sum_j w_2(\vec{x}_i, \vec{x}_j) - \sum_a w_2(\vec{x}_i, \vec{x}_a)\right]\end{aligned} \tag{18}$$

The latter two terms differ only through a scaling and whether the sums range over the observed galaxies or the random catalogue; in the absence of clustering they will be equal and cancel. Hence, they will differ only by another factor of w_2; to this order in w_2, they therefore vanish. Thus we expect no linear correction to the simple estimate of $n = N/V$.

C. First Order Solution for $w_2(\theta)$

Now, we find the maximum likelihood solution for w_2; we will defer most of the algebraic details to Appendix A.

First, we will need the derivatives of the correlation function integrals:

$$\frac{\partial W_0}{\partial w_2(\theta)} = \frac{n^2V^2}{2K^2} 2\sum_{i<j} \Theta^\theta_{i,j} \tag{19}$$

where $\Theta^\theta_{i,j}$ is one if the distance [in either angular space or real space depending on whether or not the survey has redshifts] between cells i and j lies in the bin θ. Using the definition of Landy & Szalay, this becomes

$$\frac{\partial W_0}{\partial w_2(\theta)} \equiv \frac{n^2V^2}{2} G_p(\theta) = \frac{n^2V^2 RR}{N_R^2} \tag{20}$$

where RR is the count of pairs at separation θ in the random catalogue, as in LS, and N_R is the number of random galaxies put in; this just normalizes things. Similarly,

$$\frac{\partial W_1(x_a)}{\partial w_2(\theta)} = -\frac{nV}{K}\sum_i \Theta^\theta_{i,a} \qquad \frac{\partial W_2(\vec{x}_a, \vec{x}_b)}{\partial w_2(\theta)} = \Theta^\theta_{a,b} \tag{21}$$

Another way of viewing the sums over i, j is to think of them as sums over galaxies in a random catalogue with the same geometry (and selection function) as the actual survey but with K galaxies instead of N.

We also define the data-data and data-random pair counts as

$$DD = \frac{1}{2}\sum_a \sum_b^a \Theta_{a,b}; \qquad DR = \sum_{aj} \Theta_{a,j} \tag{22}$$

where again sums over $a, b, \ldots$ are over observed galaxies and $i, j, \ldots$ are over cells or the random catalogue.

After quite a bit of algebra (see Appendix A for details), we find that

$$\frac{\partial \ln P}{\partial w_2(\theta)} = \frac{n^2 V^2}{N_R^2} RR - \frac{nV}{N_R} DR + DD - (DD) w_2(\theta) + E \tag{23}$$

where

$$\begin{aligned} E \;\; &\equiv \;\; -\frac{nV}{K}\sum_{j,a} w_2(\vec{x}_a, \vec{x}_j)\left[\frac{nV}{K}\sum_i \Theta^\theta_{i,a} - \sum_b^a \Theta^\theta_{a,b}\right] \\ &+ \;\; \sum_a \sum_b^a w_2(\vec{x}_a, \vec{x}_b)\left[\frac{nV}{K}\sum_i \Theta^\theta_{i,a} - \sum_c^b \Theta^\theta_{a,c}\right] \end{aligned} \tag{24}$$

We reiterate that the sums over a, b indices are over galaxies in the catalogue, while those over i, j are over cells or equivalently over galaxies in a random catalogue with the same geometry. In the latter case the number of cells K can be set to the number of galaxies in the random catalogue N_R.

Equation 23 would lead to a complicated (albeit linear) matrix equation for w_2. To simplify, we note that E is usually small, and is, in fact, negligible to this order in w_2, by an argument similar to that after Eq. 18. To see that this is so, note that $\sum_b \Theta^\theta_{a,b}$ is simply the number of galaxies within the bin θ surrounding the galaxy at $\vec{x}_a$. Equivalently, it is N times the fraction of galaxies in the bin θ around $\vec{x}_a$. If the galaxies were distributed randomly, this fraction would simply be $(1/K)\sum_i \Theta^\theta_{i,a}$. So the difference between the two terms in square brackets is due solely to the non-randomness of the survey, and so is proportional to w. Thus, each of the terms in square brackets in Eq. 24 are of order w_2; since they multiply terms of order w_2, these terms are quadratic in w_2. Therefore, they do not contribute at the level we are interested in.

Finally we are left with the relatively simple expression:

$$\frac{\partial \ln P}{\partial w_2(\theta)} = \frac{n^2 V^2}{N_R^2} RR(\theta) - \frac{nV}{N_R} DR(\theta) + DD(\theta) - DD(\theta) w_2(\theta). \tag{25}$$

The likelihood function therefore peaks when

$$w_2(\theta) = \frac{(n^2 V^2 / N_R^2) RR(\theta) - (nV/N_R) DR(\theta) + DD(\theta)}{DD(\theta)}. \tag{26}$$

This expression, while suggestive, is not yet complete. This expression for w_2 depends on the as yet unknown parameter n*. As we have seen in the previous subsection, the likelihood function peaks when $n = N/V$, irrespective of $w(\theta)$; hence we can *simultaneously* maximize the likelihood for both the density and correlation function. Inserting this value of n here, we find that

$$w_2(\theta) = \frac{(N^2/N_R^2)RR(\theta) - (N/N_R)DR(\theta) + DD(\theta)}{DD(\theta)}. \tag{27}$$

This then is the maximum likelihood estimator for w_2. It differs only slightly from the estimator proposed by Landy & Szalay; their estimator had RR in the denominator instead of DD. As a point of notation, we mention that LS refer to their estimator as $(DD - 2DR + RR)/RR$; the factor of 2 results from a normalization choice they have made (in the definition of d in their Eq. 46).

In §IV, we will explore these estimators in greater detail.

D. Width of Likelihood Function

We now want to calculate the width of the likelihood function. Specifically, we are interested in the matrix

$$C^{-1}_{\alpha,\beta} \equiv -\frac{\partial^2 \ln P}{\partial\lambda_\alpha \partial\lambda_\beta} \tag{28}$$

where the parameters in the theory, λ_α are the binned $w_2(\theta)$ and n. The variances for each of the individual quantities are then $1/C^{-1}_{\alpha,\alpha}$ if all other parameters are held fixed, while $C_{\alpha,\beta}$ is the general covariance matrix if all parameters are allowed to vary.

Let us first calculate $C^{-1}_{n,n}$ by differentiating Eq. 18. We find

$$C^{-1}_{n,n} = \frac{N}{n^2} - \frac{V^2}{K^2}\sum_{i,j} w_2(\vec{x}_i, \vec{x}_j). \tag{29}$$

So the variance in n is increased by the presence of correlations.

The next element is obtained by differentiating Eq. 18 with respect to w_2. We find

$$C^{-1}_{n,w_2(\theta)} = 2nV^2\frac{RR}{N_R^2} - \frac{V}{N_R}DR \tag{30}$$

*Had we kept the terms in E in our expression for w, then the estimate for $w(\theta)$ would depend on w in all other angular bins as well, i.e. on the other free parameters in the theory. After dispensing with E, we find that the estimate for $w(\theta)$ still depends on one of the other free parameters in the theory, the density n.

Note that this vanishes if there are no correlations. Note also that since Eq. 18 was accurate only up to order w_2, this expression is accurate only up to order w_2^0. This is true for $C^{-1}_{w_2,w_2}$ as well: since our expression for $\partial \ln P/\partial w_2$ is accurate only up to order w_2, the second derivative of will not contain any information about the w_2 dependence of the width. We will not be able to distinguish between the width due to a random catalogue from that due to a correlated one. Carrying through this differentiation on Eq. 25, we find

$$C^{-1}_{w_2(\theta_1),w_2(\theta_2)} = DD(\theta_1)\delta_{\theta 1,\theta 2} \tag{31}$$

where $\delta_{\theta 1,\theta 2}$ is the Kronecker delta. This agrees with Landy & Szalay's calculation, neglecting corrections of order w_2. Correlations between w_2's at different θ's would appear in the higher order terms. To assess the relative effectiveness of these two estimators, we could go back to our expansion and attempt to extract the order w_2^3 terms. Alternatively, we could perform a Monte Carlo to see which has the lower variance. We choose the latter option.

IV. Estimators of the Two Point Function

Here we would like to test the effectiveness of various estimators. Before presenting our results, we briefly review previous work. Landy & Szalay analyzed random catalogues, attempting to measure the variance of their estimator and the more traditional DD/DR. They found that the variance of their estimator was very close to the expected Poisson value:

$$\sigma^2_{\text{Poisson}} = \frac{N_R^2}{N^2 RR(\theta)}. \tag{32}$$

Note that this is simply one over the expected number of galaxy pairs per bin squared, $1/N^2_{\text{pairs}}$ On the other hand, the DD/DR estimator gave a larger variance than this at large angles. They attributed this to the fact that at large distances, the number of galaxy pairs per bin goes up and the DD/DR variance has an additional term beyond Poisson which goes as $1/N_{\text{pairs}}$. As N_{pairs} gets larger, this additional term eventually dominates. Bernstein (1993) simulated catalogues with non-zero n-point correlations. He found that the Landy & Szalay estimator had a larger variance in this correlated case than what one would expect based on Poisson-counting. As we will see, our work agrees with both of these results.

We work in a 32^2 box (the units are irrelevant). An outline of our recipe is:

1. Generate a galaxy catalogue with of order 1500 galaxies.
2. Compute the expected value of w_2 from this catalogue using three different estimators: Landy-Szalay (LS), DD/DR, and our maximum likelihood (ML).
3. Repeat steps 1 & 2, 200 times.
4. Calculate the mean and variance of these estimators over the ensemble of realizations.

To generate a galaxy catalogue (step 1), we input our desired $w_2(\theta)$, and Fourier transformed to get the power spectrum. We then used this power spectrum to generate a density field everywhere on a 32^2 grid. We chose the amplitude of w_2 to be small enough so that δ was never less than -1 anywhere on the grid. Then, we used the density field to produce a Poisson realization with at most ten galaxies per cell. Then we randomized the positions within each cell. Several comments about this proceedure: First, we are by necessity limited to small w_2. Second, we do not trust our results on scales smaller than one cell size ($x = 1$). Third, because of periodicity, w_2 obtained in this way is symmetric about the half-way point ($x = 16$), so we restrict our analysis to $x < 16$.

To calculate the expected value of w_2 (step 2), we generate thirty random catalogues with one thousand particles each.[†] We calculate DD, DR, RR in each of seven bins, the lowest at $1 < x < 3$ and the highest at $13 < x < 15$. From these, we construct the three estimators of interests.

Our results are shown in Figures 1 and 2. Figure 1 shows that all of the estimators come very close to the true value of w_2. The variance of the LS estimator is indeed significantly smaller than DD/DR at large separations when the $1/N_{\rm pairs}$ factor begins to overtake the Poisson variance. The variance of the ML estimator is similar to LS, but as shown in the top panel of Figure 1, appears to be smaller when w_2 is non-negligible.

There is one other feature of our analysis which bears note. Figure 2 shows the variance of the LS estimator as compared with the "expected" Poisson variance, Eq. 32. The variance of the LS estimator is significantly larger when w_2 is non-negligible. We believe this is real and that there are two non-negligible additional terms in the LS

[†] This should be equivalent to one catalogue of $30,000$ particles but is computationally faster to analyze. We have checked that this is sufficient by calculating the expected variance of w_2 for a random catalogue à la Landy & Szalay. We agree with the expected variance in that case.

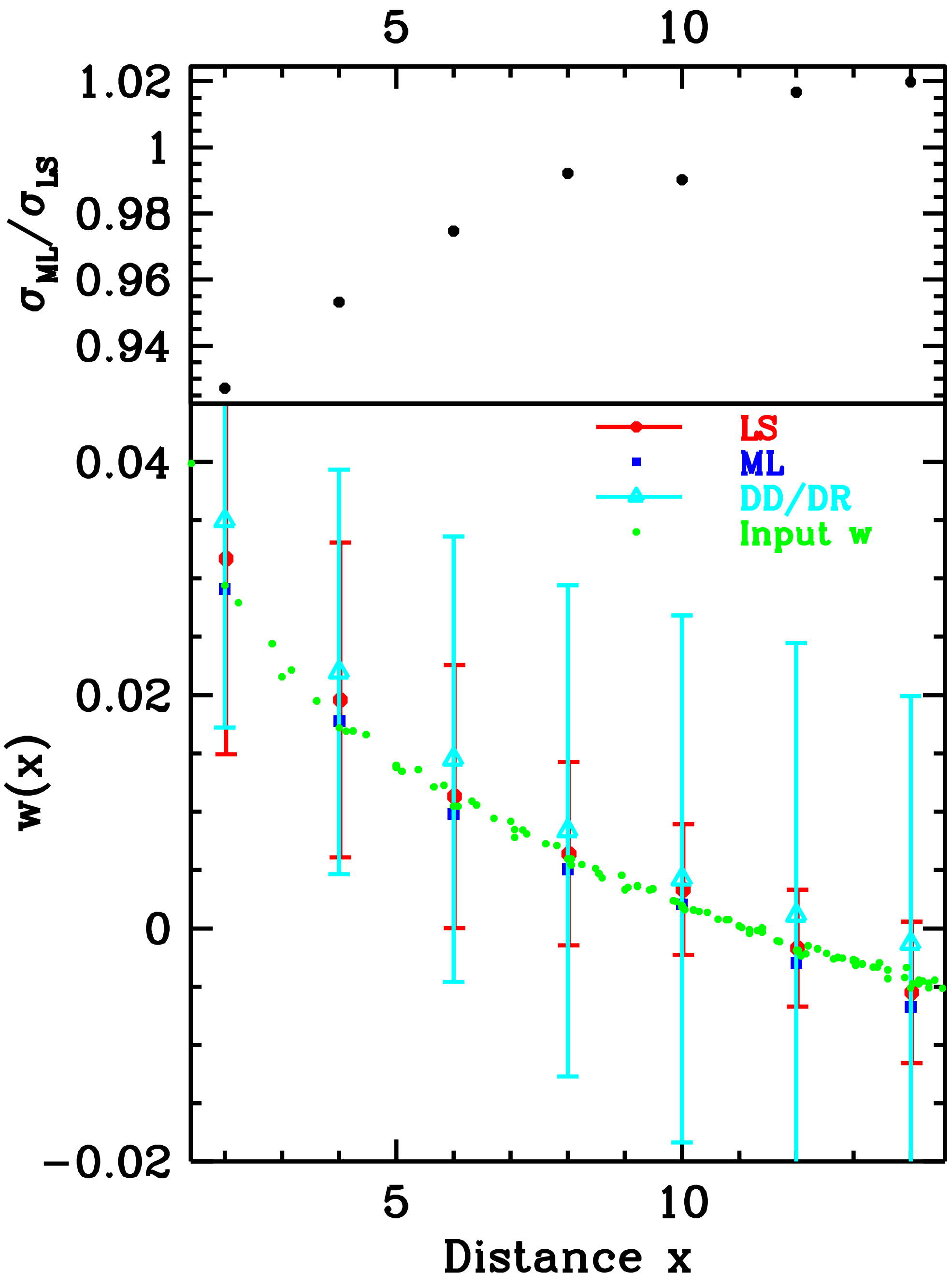

Fig. 1: The bottom panel shows the input w_2 and the estimated values, using three different estimators. The error bars for the Landy & Szalay estimator are plotted along with those for DD/DR. At large distance (with many galaxy pairs per bin) the Landy & Szalay estimator has significantly smaller variance than does DD/DR. The error bars for our maximum likelihood estimator are similar to those of the Landy & Szalay estimator so are not shown explicitly in the bottom panel. The ratio of the two is shown in the top panel. For bins with non-negligible w_2, the maximum likelihood estimator appears to have a smaller variance.

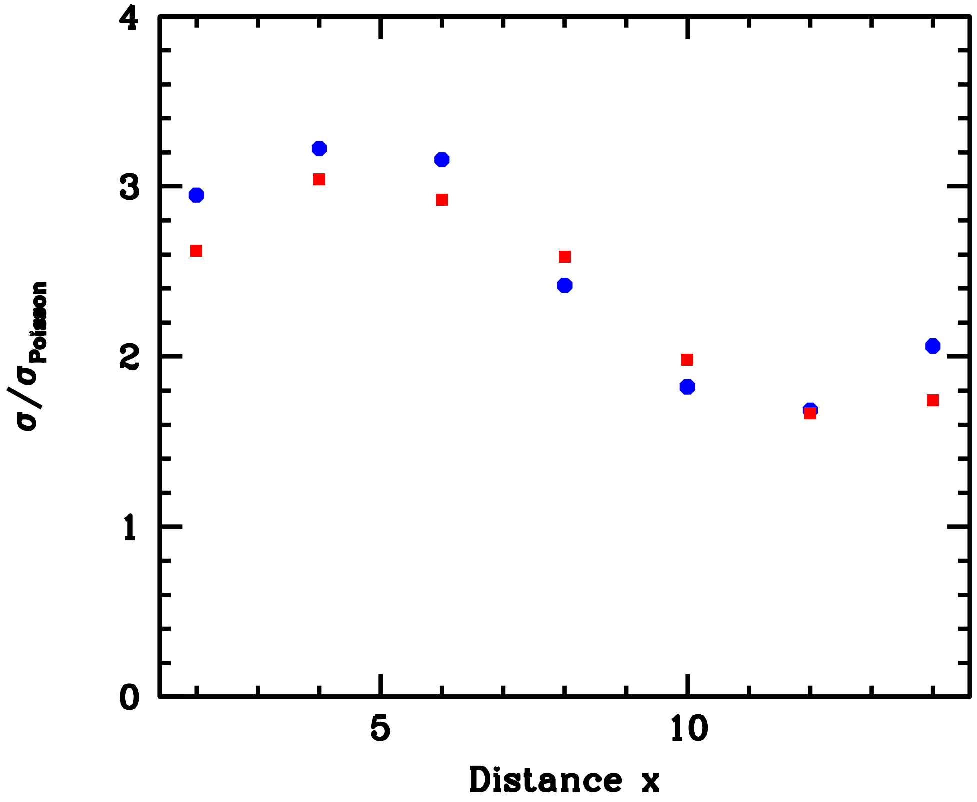

Fig. 2: The ratio of the rms variation of the Landy & Szalay estimator to the "expected" Poisson rms, $\sqrt{N_R^2/(N^2RR)}$. When w_2 is non-negligible (at small x) the rms variation is significantly larger than Poisson. (Blue) circles are variance obtained from 200 simulated catalogues; (red) squares are the variance expected from analytic results in Appendix B.

variance, one proportional to w_2/N, and one to w_2^2. Appendix B derives these additional terms analytically. Both the numerical and the analytic result agree with Bernstein's results. The differences are: we have included edge effects and we have isolated the fact that the discrepency is due to these added terms in the variance. It is not due to higher order cosmic correlations, as these are excluded by construction in our mock catalogue.

V. Conclusions and the Future

The likelihood function would be a wonderful object to compute for a galaxy survey. By construction, it would use all of the information from the survey and allow us to compare different theories. As such, we feel it is very important to explore the possibility of computing this function. Direct computation is impossible, but we have developed an approximation scheme which appears to work very well. Before specifically detailing the use we have made of this approximation, we want to emphasize that there are many ways to branch out from here:

- Use a theory such as Cold Dark Matter with its one or two free parameters. The approximation scheme developed here can then be used to extract the best fit values of these parameters.

- Generalize the approximation to include higher-order correlations. This would have the benefit of using the higher-order correlations together with the two-point function to constrain theories. Alternately, we could use the ansatz of hierarchical clustering and results from gravitational perturbation theory to generate higher-order moments from the two-point function.

- Generalize this work to Fourier space. Theories are most easily compared in Fourier space so this is a natural way to go. Just as the ML procedure generates a (diagonal) matrix equation for w_2, in Fourier space we have an integral equation for $P(k)$.

- Find a graphical method which simplifies, and helps organize, the expansion we have introduced. We have relied on arguments like: certain terms, such as E (Eq. 24), are implicitly of the order of ${w_2}^2$, even though they do not appear explicitly so. It would be nice to make these more precise and systematic.

- Go beyond the perturbative approach and try to learn something from the full likelihood function. This is not as impossible as we made it sound in §II: without actually computing $\mathcal{L}$, one might still make some very general statements.

We have made progress in the latter four areas. This will be presented in a future paper.

In this paper we have limited ourselves to one application: finding the place where the likelihood function peaks if the theoretical parameters are the binned values of the two-point function. Equivalently, we have come up with a new estimator for w_2. We found that

- The maximum likelihood (ML) estimator appears to have a slightly smaller variance than the Landy & Szalay (LS) estimator and certainly than DD/DR. To the extent that the the ML and LS estimators are similar (and they are *very* similar) this whole treatment can be thought of as further motivation for the LS estimator.
- There are additional terms in the variance of the LS estimator beyond the Poisson variance. These terms begin to dominate when correlations are non-negligible and the number of pairs of galaxies per bin is large.

This work was supported in part by DOE and NASA grant NAG5–2788 at Fermilab. We thank Istvan Szapudi for useful conversations.

[1] G. M. Bernstein, Astrophysical Journal **424**, 569 (1994).

[2] A. J. S. Hamilton, Astrophysical Journal **417**, 19 (1993).

[3] A. J. S. Hamilton, Monthly Notices of Royal Astronomical Society **289**, 285 (1997).

[4] H.C. Hewett, Monthly Notices of Royal Astronomical Society **201**, 867 (1982).

[5] S. D. Landy & A. S. Szalay, Astrophysical Journal **412**, 64 (1993).

[6] H. J. Mo, Y.P. Jing, & G. Börner, Astrophysical Journal **392**, 452 (1992).

[7] P.J.E. Peebles, The Large Scale Structure of the Universe (Princeton: Princeton University Press) (1980).

[8] I. Szapudi & A. S. Szalay, astro-ph/9704241 (1997).

[9] S. D. M. White, Monthly Notices of Royal Astronomical Society **186**, 145 (1979).

Appendix A. Derivation of Weak Correlation Limit

We will take derivatives of Eq. 11 with respect to w_2; and then expand it out to first order in w_2.

$$\begin{aligned}
\frac{\partial \ln P}{\partial w_2(\theta)} &= \frac{n^2V^2RR}{N_R^2} \\
&+ \left[\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} W_2(\vec{x}_a, \vec{x}_b) \prod_c^{a,b} W_1(\vec{x}_c)\right]^{-1} \\
&\times \left[-\frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} \prod_a^{b} W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} \frac{\partial}{\partial w_2(\theta)}\Big\{W_2(\vec{x}_a, \vec{x}_b)\prod_c^{a,b} W_1(\vec{x}_c)\Big\}\right. \\
&+ \left.\frac{1}{2^3}\sum_a \sum_b^{a} \sum_c^{a,b} \sum_d^{a,b,c} \frac{\partial}{\partial w_2(\theta)}\Big\{W_2(\vec{x}_a, \vec{x}_b)W_2(\vec{x}_c, \vec{x}_d)\prod_e^{a,b,c,d} W_1(\vec{x}_e)\Big\}\right]
\end{aligned} \tag{A1}$$

Since we are keeping only first order terms here, we have dropped the W_2^2 term in the denominator. We can go further though. The term in the denominator linear in W_2 is multiplied by the product of W_1's; these can all be set to one. Similarly, the derivative operator acting on the last term only affects the W_2's; the W_1's can again all be set to one. This gives:

$$\begin{aligned}
\frac{\partial \ln P}{\partial w_2(\theta)} &= \frac{n^2V^2RR}{N_R^2} \\
&+ \left[\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} W_2(\vec{x}_a, \vec{x}_b)\right]^{-1} \\
&\times \left[-\frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} \prod_a^{b} W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} \frac{\partial}{\partial w_2(\theta)}\Big\{W_2(\vec{x}_a, \vec{x}_b)\prod_c^{a,b} W_1(\vec{x}_c)\Big\}\right. \\
&+ \left.\frac{1}{2^3}\sum_a \sum_b^{a} \sum_c^{a,b} \sum_d^{a,b,c} \frac{\partial}{\partial w_2(\theta)}\Big\{W_2(\vec{x}_a, \vec{x}_b)W_2(\vec{x}_c, \vec{x}_d)\Big\}\right]
\end{aligned} \tag{A2}$$

Now carry out the derivatives:

$$\begin{aligned}
\frac{\partial \ln P}{\partial w_2(\theta)} &= \frac{n^2V^2RR}{N_R^2} \\
&+ \left[\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} W_2(\vec{x}_a, \vec{x}_b)\right]^{-1} \\
&\times \left[-\frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} \prod_a^{b} W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^{a} \Big(\Theta^\theta_{a,b}\prod_c^{a,b} W_1(\vec{x}_c)\right. \\
&- \frac{nV}{K} w_2(\vec{x}_a, \vec{x}_b)\sum_e^{a,b}\sum_i \Theta^\theta_{i,e}\Big)
\end{aligned}$$

$$+ \quad \frac{1}{2^3}\sum_a \sum_b^a \sum_c^{a,b} \sum_d^{a,b,c} \left(\Theta^\theta_{a,b} w_2(\vec{x}_c,\vec{x}_d) + \Theta^\theta_{c,d} w_2(\vec{x}_a,\vec{x}_b)\right)\Bigg] \tag{A3}$$

Now expand the denominator:

$$\frac{1}{\prod_a W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^a W_2(\vec{x}_a,\vec{x}_b)} = 1 - \frac{1}{2}\sum_a \sum_b^a w_2(\vec{x}_a,\vec{x}_b) - \sum_a [W_1(\vec{x}_a) - 1] \tag{A4}$$

So,

$$\begin{aligned}
\frac{\partial \ln P}{\partial w_2(\theta)} \quad & = \frac{n^2V^2}{N_R^2} RR \\
+ \quad & \left[1 - \frac{1}{2}\sum_a \sum_b^a w_2(\vec{x}_a,\vec{x}_b) - \sum_a (W_1(\vec{x}_a) - 1)\right] \\
\times \quad & \Bigg[- \frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} \prod_a^b W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^a \Big(\Theta^\theta_{a,b} \prod_c^{a,b} W_1(\vec{x}_c) \\
& - \frac{nV}{K} w_2(\vec{x}_a,\vec{x}_b) \sum_e^{a,b} \sum_i \Theta^\theta_{i,e}\Big) \\
+ \quad & \frac{1}{2^3}\sum_a \sum_b^a \sum_c^{a,b} \sum_d^{a,b,c} \left(\Theta^\theta_{a,b} w_2(\vec{x}_c,\vec{x}_d) + \Theta^\theta_{c,d} w_2(\vec{x}_a,\vec{x}_b)\right)\Bigg]
\end{aligned} \tag{A5}$$

Multiplying through, we find

$$\begin{aligned}
\frac{\partial \ln P}{\partial w_2(\theta)} \quad & = \frac{n^2V^2}{N_R^2} RR - \frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} \prod_a^b W_1(\vec{x}_a) + \frac{1}{2}\sum_a \sum_b^a \Big(\Theta^\theta_{a,b} \prod_c^{a,b} W_1(\vec{x}_c) \\
& - \frac{nV}{K} w_2(\vec{x}_a,\vec{x}_b) \sum_e^{a,b} \sum_i \Theta^\theta_{i,e}\Big) \\
+ \quad & \frac{1}{2^3}\sum_a \sum_b^a \sum_c^{a,b} \sum_d^{a,b,c} \left(\Theta^\theta_{a,b} w_2(\vec{x}_c,\vec{x}_d) + \Theta^\theta_{c,d} w_2(\vec{x}_a,\vec{x}_b)\right) \\
& - \left[\frac{1}{2}\sum_a \sum_b^a w_2(\vec{x}_a,\vec{x}_b) + \sum_a (W_1(\vec{x}_a) - 1)\right] \\
& \times \left[- \frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} + \frac{1}{2}\sum_a \sum_b^a \Theta^\theta_{a,b}\right]
\end{aligned} \tag{A6}$$

There are three sets of terms here, those which have w_2 explicitly in them; those which are independent of w_2; and those which depend on w_2 only through $W_1 - 1$. Let us treat each of these in turn.

First consider the terms independent of w_2 in Eq. A6:

$$\frac{n^2V^2}{N_R^2} RR - \frac{nV}{K}\sum_{b,i} \Theta^\theta_{i,b} + \frac{1}{2}\sum_a \sum_b^a \Theta^\theta_{a,b} = \frac{n^2V^2}{N_R^2} RR - \frac{nV}{N_R} DR + DD \tag{A7}$$

Next consider the terms which depend explicitly on w_2.

$$\begin{aligned}
&-\frac{nV}{2K}\sum_a\sum_b^{a} w_2(\vec{x}_a,\vec{x}_b)\sum_e^{a,b}\sum_i \Theta^\theta_{i,e} \quad + \quad \frac{1}{2^3}\sum_a\sum_b^{a}\sum_c^{a,b}\sum_d^{a,b,c}\Big(\Theta^\theta_{a,b}w_2(\vec{x}_c,\vec{x}_d)+\Theta^\theta_{c,d}w_2(\vec{x}_a,\vec{x}_b)\Big)\\
&-\frac{1}{2}\sum_a\sum_b^{a} w_2(\vec{x}_a,\vec{x}_b)\Big[-\frac{nV}{N_R}DR \quad + \quad DD\Big]
\end{aligned} \tag{A8}$$

We claim that the two terms in the quadruple sum on the first line are identical. To see this, first switch indices $a \leftrightarrow c$; $b \leftrightarrow d$ in the first term. Then it is:

$$\sum_c\sum_d^{c}\sum_a^{c,d}\sum_b^{c,d,a}\Theta^\theta_{c,d}w_2(\vec{x}_a,\vec{x}_b) = \sum_a\sum_c^{a}\sum_d^{a,c}\sum_b^{c,d,a}\Theta^\theta_{c,d}w_2(\vec{x}_a,\vec{x}_b) \tag{A9}$$

where we have switched the summations and taken care to guard against summing over identical $a = c$ for example. Continuing in this fashion, we get the second term in the quadruple sum. Thus, all terms with explicit w_2's in them are:

$$\sum_a\sum_b^{a} w_2(\vec{x}_a,\vec{x}_b)\left[-\frac{nV}{2K}\sum_e^{a,b}\sum_i\Theta^\theta_{i,e} + \frac{1}{2^2}\sum_c^{a,b}\sum_d^{a,b,c}\Theta^\theta_{c,d} - \frac{1}{2}\Big(-\frac{nV}{N_R}DR + DD\Big)\right] \tag{A10}$$

Consider the first term in brackets. We can rewrite the sum over e as

$$\sum_e^{a,b} = \sum_e(1-\delta_{ae})(1-\delta_{be}) = \sum_e(1-\delta_{ae}-\delta_{be}) \tag{A11}$$

Note that since a is never equal to b, this is exactly true. Now the unrestricted sum over e simply gives DR with the sign and coefficient exactly right to cancel the third term in brackets. So the terms with explicit w_2 dependence are:

$$\sum_a\sum_b^{a} w_2(\vec{x}_a,\vec{x}_b)\left[\frac{nV}{2K}\sum_i(\Theta^\theta_{i,a}+\Theta^\theta_{i,b}) + \frac{1}{2^2}\sum_c^{a,b}\sum_d^{a,b,c}\Theta^\theta_{c,d} - \frac{1}{2}DD\right] \tag{A12}$$

The same argument holds for the DD terms. The only terms which remain in the sums over c, d are those wherein c or d equals a, b, c. Let us do this carefully because there might be subtleties here.

$$\begin{aligned}
\sum_c^{a,b}\sum_d^{a,b,c} &= \sum_c^{a,b}\sum_d^{c}(1-\delta_{da}-\delta_{db})\\
&= \sum_c\sum_d^{c}[1-(\delta_{ac}+\delta_{bc})] - \sum_c^{a,b}\sum_d(\delta_{da}+\delta_{db})
\end{aligned} \tag{A13}$$

So the explicit w_2 terms become

$$\sum_a \sum_b^a w_2(\vec{x}_a, \vec{x}_b) \left[\frac{nV}{2K} \sum_i (\Theta^\theta_{i,a} + \Theta^\theta_{i,b}) - \frac{1}{2^2} [\sum_d^a \Theta^\theta_{a,d} + \sum_d^b \Theta^\theta_{b,d} + \sum_c^{a,b} (\Theta^\theta_{c,a} + \Theta^\theta_{b,c})] \right] \tag{A14}$$

We want to focus on one of these terms: the one in which the index d is equal to a or b. Thus rewrite this last line as

$$\sum_a \sum_b^a w_2(\vec{x}_a, \vec{x}_b) \left[\frac{nV}{K} \sum_i \Theta^\theta_{i,a} - \sum_d^b \Theta^\theta_{a,d} \right] - \frac{1}{2} \sum_a \sum_b^a w_2(\vec{x}_a, \vec{x}_b) \Theta^\theta_{a,b} \tag{A15}$$

where we have made use of the symmetry between a, b to sum up lots of identical terms. Now consider the last term. The $\Theta^\theta_{a,b}$ requires all separations between $\vec{x}_a$ and $\vec{x}_b$ to lie within the bin θ. Within this bin, by definition, w_2 is constant [this is the theory we are trying to solve for]. Thus, w_2 comes out of the sum and we are left with simply DD. So the terms with only explicit w_2 dependence are

$$\sum_a \sum_b^a w_2(\vec{x}_a, \vec{x}_b) \left[\frac{nV}{K} \sum_i \Theta^\theta_{i,a} - \sum_d^b \Theta^\theta_{a,d} \right] - (DD) w_2(\theta) \tag{A16}$$

We can now reinsert all this back into Eq. A6.

$$\begin{aligned} \frac{\partial \ln P}{\partial w_2(\theta)} &= \frac{n^2 V^2}{N_R^2} RR - \frac{nV}{N_R} DR + DD - (DD) w_2(\theta) \\ &- \frac{nV}{K} \sum_{b,i} \Theta^\theta_{i,b} \sum_a^b [W_1(\vec{x}_a) - 1] + \frac{1}{2} \sum_a \sum_b^a \Theta^\theta_{a,b} \sum_c^{a,b} [W_1(\vec{x}_c) - 1] \\ &- \sum_c (W_1(\vec{x}_c) - 1) \Bigg[- \frac{nV}{K} \sum_{b,i} \Theta^\theta_{i,b} + \frac{1}{2} \sum_a \sum_b^a \Theta^\theta_{a,b} \Bigg] \\ &+ \sum_a \sum_b^a w_2(\vec{x}_a, \vec{x}_b) \left[\frac{nV}{K} \sum_i \Theta^\theta_{i,a} - \sum_d^b \Theta^\theta_{a,d} \right] \end{aligned} \tag{A17}$$

The terms in the middle two lines nearly cancel, except for the restrictions on the sums. The only terms that remain from these two lines are those in which c on the third line equals a or b. Thus we are lead directly to Eq. 23.

Appendix B. Variance of the Landy & Szalay Estimator

To derive the variance of the Landy & Szalay estimator, let us rewrite it in the following form:

$$w_2^{\mathrm{LS}}(\theta) = \sum_{i,j} W_{i,j}(\theta)(q_i - \alpha q_i^r)(q_j - \alpha q_j^r)\,, \tag{B1}$$

where

$$W_{i,j}(\theta) = \frac{\Theta_{i,j}(\theta)}{\sum_{m,n}\Theta_{m,n}(\theta)\alpha^2 q_m^r q_n^r}\,, \tag{B2}$$

and q_i equals one where the cell contains a galaxy, and equals zero otherwise, whereas q_i^r is similarly defined for the random catalogue. The factor α scales the average density in random catalogue back to the level in the actual catalogue: in other words, $\alpha = N/N_R$ with N and N_R being the total number of galaxies in the actual catalogue and in the random catalogue respectively.

It can be shown that the introduction of the random catalogue introduces extra variance (i.e. variance of the random catalogue itself), which could be eliminated if one uses a large number of random catalogues, or if one allows the number of galaxies in random catalogue to increase dramatically. We will compute the variance of $w_2^{\mathrm{LS}}(\theta)$ in this limit, in which case, one can replace every αq_i^r by $\bar{q}$ where $\bar{q} = \langle q_i \rangle$. Hence, Eq. B1 is equivalent to

$$w_2^{\mathrm{LS}}(\theta) = \sum_{i,j} \tilde{W}_{i,j}(\theta)\delta_i\delta_j \tag{B3}$$

with

$$\tilde{W}_{i,j}(\theta) = \frac{\Theta_{i,j}(\theta)}{\sum_{m,n}\Theta_{m,n}(\theta)} \tag{B4}$$

and $\delta_i = (q_i - \bar{q})/\bar{q}$.

The variance is defined by $\langle [w_2^{\mathrm{LS}}(\theta)]^2\rangle - \langle w_2^{\mathrm{LS}}(\theta)\rangle^2$. This means one needs the following quantity:

$$\begin{aligned} &\langle \delta_i\delta_j\delta_k\delta_l\rangle - \langle\delta_i\delta_j\rangle\langle\delta_k\delta_l\rangle \\ &= [w_2(\vec{x}_i, \vec{x}_k) + \delta_{ik}\bar{q}_i^{-1}][w_2(\vec{x}_j, \vec{x}_l) + \delta_{jl}\bar{q}_i^{-1}] + (k \leftrightarrow l) \end{aligned} \tag{B5}$$

which is adopted from Hamilton (1997). This expression holds in the small pixel (continuum) limit, with the restriction $i \neq j$ and $k \neq l$ (for nonzero θ) and for Gaussian random underlying field. We have also ignored terms the contributions of which to the variance are of order of $1/N$ smaller than those we have kept.

Putting everything together, we obtain:

$$\begin{aligned} \text{Variance} \quad &= \frac{2}{\bar{q}^2\sum_{i,j}\Theta_{i,j}(\theta)} + \frac{4\sum_{i,j,l}\Theta_{i,j}(\theta)\Theta_{i,l}(\theta)w_2(\vec{x}_j,\vec{x}_l)}{\bar{q}[\sum_{m,n}\Theta_{m,n}(\theta)]^2} \\ &+ \frac{2\sum_{i,j,k,l}\Theta_{i,j}(\theta)\Theta_{k,l}(\theta)w_2(\vec{x}_i,\vec{x}_k)w_2(\vec{x}_j,\vec{x}_l)}{[\sum_{m,n}\Theta_{m,n}(\theta)]^2} \end{aligned} \tag{B6}$$

The first term on the right is the Poisson variance given by Landy & Szalay (1993). The second two terms arise because of finite two-point correlations. They have been derived before by Bernstein (1993), ignoring edge effects, and by Mo, Jing, & Börner, ignoring discreteness effects. Note that both Landy & Szalay and Bernstein obtained extra terms for the Poisson variance, which we have ignored because they are of the order of $1/N$ smaller than those we have kept. The variance in Eq. B6 can be estimated using the standard technique of counting pairs, triplets and quadruplets using a random catalogue with the same geometry. For instance one can estimate $\bar{q}^2 \sum_{i,j} \Theta_{i,j}(\theta)/2$ by $[N^2/N_R^2] RR(\theta)$.